This paper is published as part of a PCCP Themed Issue on:

Nucleic Acid Simulations

Guest Editors: Modesto Orozco and Charles Laughton

Editorial

Nucleic Acid Simulations
Modesto Orozco and Charles Laughton, *Phys. Chem. Chem. Phys.*, 2009
DOI: 10.1039/b921472k

Perspectives

Viral assembly: a molecular modeling perspective
Stephen C. Harvey, Anton S. Petrov, Batsal Devkota and Mustafa Burak Boz, *Phys. Chem. Chem. Phys.*, 2009
DOI: 10.1039/b912884k

Simulation of DNA catenanes
Alexander Vologodskii and Valentin V. Rybenkov, *Phys. Chem. Chem. Phys.*, 2009
DOI: 10.1039/b910812b

Papers

Stabilization of radical anion states of nucleobases in DNA
Alexander A. Voityuk, *Phys. Chem. Chem. Phys.*, 2009
DOI: 10.1039/b910690a

Effects of the biological backbone on DNA–protein stacking interactions
Cassandra D. M. Churchill, Lex Navarro-Whyte, Lesley R. Rutledge and Stacey D. Wetmore, *Phys. Chem. Chem. Phys.*, 2009
DOI: 10.1039/b910747a

The impact of monovalent ion force field model in nucleic acids simulations
Agnes Noy, Ignacio Soteras, F. Javier Luque and Modesto Orozco, *Phys. Chem. Chem. Phys.*, 2009
DOI: 10.1039/b912067j

Structural ultrafast dynamics of macromolecules: diffraction of free DNA and effect of hydration
Milo M. Lin, Dmitry Shorokhov and Ahmed H. Zewail, *Phys. Chem. Chem. Phys.*, 2009
DOI: 10.1039/b910794k

Simulation of DNA double-strand dissociation and formation during replica-exchange molecular dynamics simulations
Srinivasaraghavan Kannan and Martin Zacharias, *Phys. Chem. Chem. Phys.*, 2009
DOI: 10.1039/b910792b

Sensors for DNA detection: theoretical investigation of the conformational properties of immobilized single-strand DNA
Vincenzo Barone, Ivo Cacelli, Alessandro Ferretti, Susanna Monti and Giacomo Prampolini, *Phys. Chem. Chem. Phys.*, 2009
DOI: 10.1039/b914386f

Relaxation dynamics of nucleosomal DNA
Sergei Y. Ponomarev, Vakhtang Putkaradze and Thomas C. Bishop, *Phys. Chem. Chem. Phys.*, 2009
DOI: 10.1039/b910937b

Dynamics of a fluorophore attached to superhelical DNA: FCS experiments simulated by Brownian dynamics
Tomasz Wocjan, Jan Krieger, Oleg Krichevsky and Jörg Langowski, *Phys. Chem. Chem. Phys.*, 2009
DOI: 10.1039/b911857h

A detailed binding free energy study of 2 : 1 ligand–DNA complex formation by experiment and simulation
Witcha Treesuwan, Kitiyaporn Wittayanarakul, Nahoum G. Anthony, Guillaume Huchet, Hasan Alniss, Supa Hannongbua, Abedawn I. Khalaf, Colin J. Suckling, John A. Parkinson and Simon P. Mackay, *Phys. Chem. Chem. Phys.*, 2009
DOI: 10.1039/b910574c

Molecular simulation of conformational transitions in biomolecules using a combination of structure-based potential and empirical valence bond theory
Giuseppe de Marco and Péter Várnai, *Phys. Chem. Chem. Phys.*, 2009
DOI: 10.1039/b917109f

Dependence of A-RNA simulations on the choice of the force field and salt strength
Ivana Beššeová, Michal Otyepka, Kamila Réblová and Jiří Šponer, *Phys. Chem. Chem. Phys.*, 2009
DOI: 10.1039/b911169g

Protein–DNA binding specificity: a grid-enabled computational approach applied to single and multiple protein assemblies
Krystyna Zakrzewska, Benjamin Bouvier, Alexis Michon, Christophe Blanchet and Richard Lavery, *Phys. Chem. Chem. Phys.*, 2009
DOI: 10.1039/b910888m

Evaluation of molecular modelling methods to predict the sequence-selectivity of DNA minor groove binding ligands
Hao Wang and Charles A. Laughton, *Phys. Chem. Chem. Phys.*, 2009
DOI: 10.1039/b911702d

Mesoscale simulations of two nucleosome-repeat length oligonucleosomes
Tamar Schlick and Ognjen Perišić, *Phys. Chem. Chem. Phys.*, 2009
DOI: 10.1039/b918629h

On the parameterization of rigid base and basepair models of DNA from molecular dynamics simulations
F. Lankaš, O. Gonzalez, L. M. Heffler, G. Stoll, M. Moakher and J. H. Maddocks, *Phys. Chem. Chem. Phys.*, 2009
DOI: 10.1039/b919565n

Charge transfer equilibria of aqueous single stranded DNA
Marco D'Abramo, Massimiliano Aschi and Andrea Amadei, *Phys. Chem. Chem. Phys.*, 2009
DOI: 10.1039/b915312h

Structural ultrafast dynamics of macromolecules: diffraction of free DNA and effect of hydration

Milo M. Lin, Dmitry Shorokhov and Ahmed H. Zewail*

Received 1st June 2009, Accepted 10th August 2009
First published as an Advance Article on the web 15th September 2009
DOI: 10.1039/b910794k

Of special interest in molecular biology is the study of structural and conformational changes which are free of the additional effects of the environment. In the present contribution, we report on the ultrafast unfolding dynamics of a large DNA macromolecular ensemble *in vacuo* for a number of temperature jumps, and make a comparison with the unfolding dynamics of the DNA in aqueous solution. A number of coarse-graining approaches, such as kinetic intermediate structure (KIS) model and ensemble-averaged radial distribution functions, are used to account for the transitional dynamics of the DNA without sacrificing the structural resolution. The studied ensembles of DNA macromolecules were generated using distributed molecular dynamics (MD) simulations, and the ensemble convergence was ensured by monitoring the ensemble-averaged radial distribution functions and KIS unfolding trajectories. Because the order–disorder transition in free DNA implies unzipping, coiling, and strand-separation processes which occur consecutively or competitively depending on the initial and final temperature of the ensemble, DNA order–disorder transition *in vacuo* cannot be described as a two-state (un)folding process.

1. Introduction

Deoxyribonucleic acid (DNA) encodes for proteins and is therefore the genetic blueprint for all life, responsible for both diversities and similarities throughout the biosphere. Because the nucleotide sequence in DNA is organism-specific, and because it does not vary noticeably from one cell to another under normal circumstances, DNA macromolecules represent an invaluable source of biological, medical and forensic information which has been subject to intense investigation since the very discovery of their double-helix structure[1] from X-ray fiber diffraction patterns[2] in 1953. With single-crystal X-ray diffraction,[3] the macromolecular architecture was resolved at the atomic level. Manipulations of DNA sequence and structure *via* direct genetic engineering are now widely used to improve crops and livestock quality,[4] as well as to produce biological tissues and substances with desired characteristics,[5] whereas the so-called bottom-up fabrication yields DNA nanostructures of different shapes (''DNA origami'').[6] The double stranded helix of the DNA is unique in many aspects, including chemical and thermodynamic stability, packing efficiency, and site-specific strand separation which prevent harmful mutations, facilitate folding, and allow for transcription, respectively. Both molecular structure and (un)folding dynamics of DNA are, therefore, central to our understanding of a variety of processes taking place *in vivo*.

DNA double helices are known to be well suited for tight packing of genetic material (notably, the DNA in a single human chromosome reaches the centimeter length scale when totally extended). Thus, there are 23 pairs of such macromolecules in the micron-sized nucleus of a healthy cell.[7] Under physiological conditions, the stability of a DNA duplex stems from a delicate balance of a number of competing forces and mechanisms.[8] In particular, the Coulomb repulsion between negatively charged phosphate groups is compensated by stacking and hydrogen bonding interactions between DNA bases and by screening effect of water and surrounding ions. Hydration affects the stability of the duplex because the change from aqueous to less polar solvents reportedly leads to pronounced conformational transitions and/or disruption of the helical pattern.[9] Similarly, it has been demonstrated by X-ray fiber diffraction that, upon variation of the relative humidity of fiber environment, the molecular structures assumed by DNA fibers vary from A-DNA to Z-DNA (Fig. 1), and that the hydration-driven transitions between the DNA conformers are fully reversible.[10] A recent determination of the structure of DNA in single crystals of nucleosome core particles revealed that the DNA is predominantly in the B-form with local distortions and irregularities, which facilitate its superhelical path in the nucleosome.[11] Generally, there seems to be little reservation about prevalence of B-DNA occurring *in vivo*, although in various situations (*e.g.*, around histones) the molecule adapts a bent configuration.[12a]

In light of the structural integrity of B-form DNA in solution, the physical source of this stability becomes a highly relevant topic. In particular, what features of the duplex are preserved *in vacuo* and are therefore ''inherent'' to the physics of DNA alone? Because the replication and transcription of DNA are dynamical processes involving strand separation, an equally important goal is the understanding of the effect of solvent on both conformational and unfolding dynamics. Although biological macromolecules often tend to undergo

Physical Biology Center for Ultrafast Science and Technology, Arthur Amos Noyes Laboratory for Chemical Physics, California Institute of Technology, Pasadena, CA 91125, USA. E-mail: zewail@caltech.edu; Fax: +1 (626) 792 8456

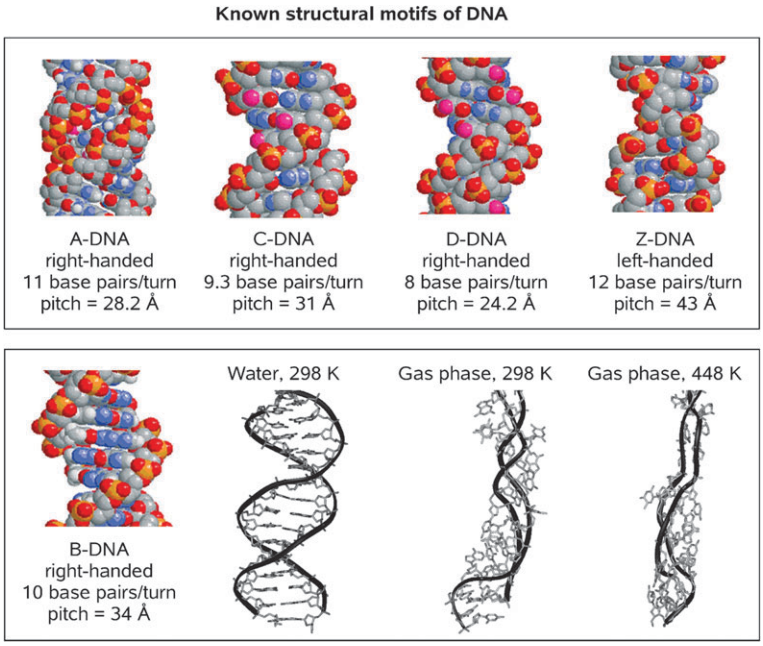

Fig. 1 Known structural motifs of DNA. Shown are known types of DNA folding in the presence of hydrating environment (A–D, Z) as obtained from X-ray fiber diffraction experiments[10,12a] and extended structure *in vacuo* as obtained recently from MD simulations.[8]

local distortions, the structures they commonly assume are relatively robust, persisting in a wide variety of environments. Thus, according to electrospray experimental observations, DNA retains its major structural features even in the absence of the hydrating water layer.[13] However, because direct experimental determination of detailed molecular structures of large, flexible biopolymers *in vacuo* and in solution is not feasible at present, theoretical methods remain the guiding force in exploring the configurational space of DNA/RNA in various environments.[14b]

Remarkably, the double-helix architecture of nucleic acids gives rise to a number of persistent structural features which can be efficiently exploited in experimental and theoretical studies. For example, by analyzing the diffraction from spatially aligned DNA fibers, both the helical symmetry and macromolecular structure of the fibers can be deduced.[12] This is typically accomplished by calculating the diffraction pattern using a "theoretical" (anticipated) structure of a single fiber which is presumed to be known with a sufficiently high resolution. By comparing the theoretical pattern with the "observed" (experimental) one, which originates from a large number of well-oriented, coherently scattering fibers, an improved structural model of the fiber may be constructed. However, of special interest in molecular biology is the study of structural and conformational changes which are free of the effects of solvation, crystallization or external ordering imposed on the specimen.[15] Because of their large size and unprecedented flexibility, DNA macromolecules possess a myriad of quasi-random structural configurations during the course of an order–disorder transition, and this complexity may, naively, suggest the masking of any significant change in diffraction. However, as we demonstrate in detail in the subsequent Sections, an accurate theoretical mapping of macromolecular ensembles which consist of hundreds of (non-interacting) DNA duplexes indicates that the pronounced features of the quasi-periodic structure (*spatial resonance*)[16] in DNA may be used as a natural measure of the disruption of the double-helix ordering in both space and time.

For over a decade, molecular dynamics (MD) simulations have been capable of reproducing the structure and dynamics of large DNA macromolecules in aqueous solutions.[17] Although MD simulations of gas-phase nucleic acids have been performed even earlier for sub-nanosecond dynamics,[18] in a recent computational study, Rueda *et al.* demonstrated for the first time that a somewhat distorted DNA duplex might be stable in the gas phase on the (sub)microsecond time scale.[8] Sample conformations of a free DNA duplex as obtained from their MD simulations carried out at room temperature and at $T = 448$ K are highlighted in Fig. 1. Importantly, the conformational transition due to vaporization of DNA should occur very rapidly given the size of the macromolecule under scrutiny (the complete equilibration is reportedly achieved on a few-nanosecond time scale). As seen in Fig. 1, the equilibrated DNA duplexes are strongly extended along the helix axis which results in local (site-specific) irregularities and somewhat shallower grooves as compared to the canonical structure of B-DNA.

However, the extended duplex structures retain many features characteristic of the canonical (hydrated) DNA configuration irrespective of both the temperature ($T \leq 448$ K) and the neutralization protocol employed in the numerical simulations. Despite the similarities which reportedly exist between the equilibrated duplexes and C-DNA, the former are better described as mechanically-stretched (elongated) double helices.[19] Interestingly, the vaporization does not have a dramatic effect on the conformational preferences of DNA nucleotides. Thus, 60 to 90% of canonical hydrogen bonds are preserved in the gas phase at 298 K. However, the fraction of

noncanonical hydrogen bonds, which is negligible in solution, increases to $\leq$40% in the gas phase. The DNA bases remained well-stacked throughout the simulations, but the stacking direction was no longer parallel to the helix axis (see Fig. 1). According to the results of ref. 8, no strand separations were observed in any of their MD trajectories.

In what follows, we discuss ultrafast structural dynamics of DNA unfolding in the gas phase as obtained from ensemble-convergent MD simulations carried out for a number of charged states ($Q = -6$ or -12) and temperature jumps ($T_0 = 300$ K; $300 \leq \Delta T \leq 1200$ K) and make a comparison with unfolding dynamics of the same macromolecule in aqueous solution. In contrast to earlier theoretical work, the focus here is on the dynamics of DNA duplex unfolding, including the influence of hydration on the mechanisms and time scales of this process. In addition to the use of native-contact metrics to measure base-pairing disruption, we also investigate the conformational dynamics of the duplex *via* the radial distribution functions of ultrafast electron diffraction (UED) simulations. Though certain apparent implications for future UED experiments are briefly outlined in the subsequent Section, the above-mentioned radial distribution functions are only invoked here as an intuitively appealing coarse-graining model. As all analyses were performed on a large number of independent unfolding trajectories, we elucidate the dramatic effect of solvation on the structural and dynamical properties of DNA in an ensemble-convergent manner. It is noteworthy that self-sufficiency and credibility of the results reported below are ensured by the data redundancy and signal-to-noise ratio comparable to those typically obtained during the course of UED experiments.

2. Energy landscapes and electron diffraction

In the present work, large-scale molecular ensembles of 5′-d(CGCGGTGTCCGCG)-3′ DNA duplexes (PDB ID: 1LAI)[20] were generated using the CHARMM[21] suite of programs both *in vacuo* and in aqueous solution, and their spatiotemporal evolution was monitored for a variety of temperature jumps during the course of distributed MD simulations. Such simulations provide a state-of-the-art theoretical account of structural interconversions within the ensemble because they are based on realistic interatomic interaction potentials which determine molecular motions at each particular point in time. With our newly built super-computer cluster, which currently features 32 dual quad-core Intel E5345 compute nodes, 12 GB RAM/node, $\sim$20 TB of network-attached disk storage, and a 1-GigE network inter-connection mesh, we are now poised to explore such dynamics of complex energy landscapes. The structural interconversions which involve numerous degrees of mechanical freedom, such as conformational changes in biological macromolecules, are easily modeled and the ensemble convergence is achieved at increasingly longer time scales. Atomic-scale spatiotemporal resolution combined with a massive statistical redundancy, which is sufficient to ensure the desired degree of ensemble convergence, constitute the basis for accuracy and reproducibility of the numerical experiments presented below.

In the earlier reports from this laboratory,[14] we outlined the importance of collapsed (but not folded) macromolecular structures in the DNA order–disorder transitions. Contrary to the textbook wisdom, it is now widely accepted that DNA (un)folding is not a two-state kinetic process because, on the ultrafast time scale, locally stable states of the DNA (un)folding trajectories are not limited to a set of completely folded and completely unfolded molecular configurations.[14] Thus, an adequate level of coarse graining is required to parameterize the configurational space of the DNA macromolecule: ideally, the model parameterization should allow for the entire variety of intermediate states, both collapsed and (partially) unfolded. Because conventional approaches to coarse graining such as, *e.g.*, percent native, or non-native, base contacts or root-mean-square deviations (RMSD) from the native structure,[22] do not provide the level of detail required to resolve the actual structures of the intermediate states, we recently employed the kinetic intermediate structure (KIS) model to parameterize the configurational space of small DNA hairpins in aqueous solution.[14b] Notably, the essential mechanisms distilled by the model are illustrative of the key insights possible with well-chosen coarse-graining models of macromolecular folding.[23]

Within the KIS model, we consider native (Watson–Crick) base pairs, and the reaction coordinates i and j are chosen to be the number of unzipped base pairs on the loop and free ends of the stem, respectively (Fig. 2). The choice of coordinates implicitly constrains the KIS model to the single-sequence approximation (SSA) which excludes all structures with internal bulges or loops (besides the hairpin loop).[24] All intermediate states are then represented by unique coordinates (i,j) on the surface, with the hairpin native state at (0,0). The only state that does not have a unique point on the landscape is the unfolded-structure ensemble, which is represented by the points on the diagonal boundary of the coordinate space (Fig. 2). Each state (i,j) corresponds to an ensemble of structures that share the same base pairing but may differ in their detailed atomic coordinates. The free-energy landscape $\Delta G(i,j)$ which defines the (un)folding behavior is then obtained by calculating the free energy for each (i,j)-state with respect to the native state of the molecule at (0,0), using the thermodynamic parameters employed by Kuznetsov *et al.*[25] Significantly, the KIS model, which can be easily extended for short DNA duplexes as discussed below, retains the comprehensive picture of kinetics without sacrificing the structural resolution. However, because the model is limited to a local (native-contact-specific) description of intermediate states, it is not suited to describe the three-dimensional conformations characterized by a (partial) strand separation.

In the present contribution we utilize the (time-dependent) *ensemble-averaged radial distribution function*, $\langle f(r, t)\rangle_n$, where n is the number of macromolecules in a numerically generated ensemble, to quantify both local and global conformational changes which accompany the order–disorder transition in DNA. Because $\langle f(r, t)\rangle_n$ provides a snapshot of the density distribution of internuclear distances r_{ij}, $i \neq j$, $i, j \leq N$, where N is the number of atoms in the macromolecule, throughout the simulated ensemble at each particular point in time t, $\Delta\langle f(r, t_m, t_k)\rangle_n = \langle f(r, t_m)\rangle_n - \langle f(r, t_k)\rangle_n$ represents a natural measure of the ensemble-averaged conformational changes

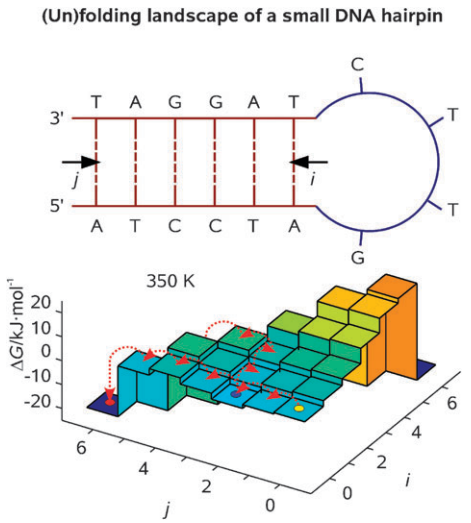

Fig. 2 (Un)folding landscape of a small DNA hairpin. Kinetic intermediate structure (KIS) model and single-sequence approximation (SSA)[14b] provide a schematic representation of the (un)folding reaction coordinates and the configurational landscape of the hairpin in aqueous solution.

which occur during the time interval $\Delta t = t_m - t_k$. The procedure of obtaining $\langle f(r, t)\rangle_n$ has been described in detail in a number of sources.[16] Briefly, the electron diffraction patterns $sM(s, t)$, where s is the magnitude of momentum transfer vector between an incident electron and an elastically scattered electron, are calculated at t_m and t_k using a locally modified version of UEDANA[26] code for each DNA macromolecule in the ensemble, and further averaged to yield $\langle sM(s, t_m)\rangle_n$ and $\langle sM(s, t_k)\rangle_n$. The resulting $\langle sM(s, t)\rangle_n$ are Fourier-transformed to obtain $\langle f(r, t_m)\rangle_n$, $\langle f(r, t_k)\rangle_n$, and $\Delta\langle f(r, t_m, t_k)\rangle_n$. Because $\langle f(r, t)\rangle_n$ and $\Delta\langle f(r, t_m, t_k)\rangle_n$ represent the ensemble-wide structural changes in real space, they have more intuitive appeal than $\langle sM(s, t)\rangle_n$ and $\Delta\langle sM(s, t_m, t_k)\rangle$, but the latter should also be examined in detail in order to assess the potential feasibility of an electron diffraction experiment.

Because the 3D coordinates of atoms in a macromolecule are projected onto a 1D momentum transfer space (Fourier, or reciprocal, space) as a superposition of damped oscillations (each of which represents a particular internuclear separation r_{ij} within the studied molecular structure), the resulting electron diffraction patterns, radial distribution functions, and their differences constitute an intuitive coarse-graining approach with which to analyze the ensemble-level temporal evolution pertinent to order–disorder transitions in DNA. Due to additivity of the scattering terms in diffraction, the ensemble-wide set of internuclear distances can be easily decomposed into relevant subsets, such as bond distances, nearest-neighbor distances, or internuclear distances constrained to specific macromolecular moieties. Because the resulting term-by-term separability of $\langle f(r, t)\rangle_n$ permits the analysis of structural changes characteristic of the chosen groups of atoms, $\Delta\langle f(r, t_m, t_k)\rangle_n$ is ideally suited for isolation of the key structural motifs, *e.g.*, those giving rise to the spatial resonance in diffraction,[16] throughout the entire macromolecular ensemble. Thus, the internuclear distance density distribution within the subset of phosphorus atoms, {P}, which occur quasi-periodically in the sugar backbones of DNA duplexes, $\langle f_P(r, t)\rangle_n$, can be used to monitor the deterioration of the global structural ordering in the ensemble (*e.g.*, strand separation and coiling) in both space and time. This degree of coarse graining is essential because (i) the distance density distribution within the subset is free from obscuring effects of incoherently-distributed internuclear distances and (ii) the KIS model, which proved to be successful in describing the base-paring contacts in DNA, does not provide an insight into the geometric conformations and orientation of its strands.

In the remainder of this Section, we illustrate the above-mentioned principles with simulated $f(r)$ and $sM(s)$ patterns of a single, isolated B-DNA macromolecule and make connections to the potential UED measurements. Radial distribution functions of the full duplex as obtained for the NMR structure of the 1LAI 13-mer found in aqueous solution[20] are depicted in Fig. 3. Also shown in the Figure are corresponding patterns of individual (uncoupled) DNA strands as they appear in the duplex. We note that the differences between the two kinds of patterns are really striking, which is indicative of a very pronounced change in diffraction upon the complete separation of the strands (the experimentally-observed difference would be even larger because the uncoupled DNA strands would also lose their structural ordering associated with both stacking and helicity characteristic of the B-DNA structure). Importantly, calculated diffraction patterns of the macromolecule and the uncoupled strands are almost indistinguishable at $s > 3$ Å^{-1} which indicates that the features arising from the interstrand r_{ij} distances are largely concentrated in the innermost area of the scattering pattern.

In order to identify the origin of the spatial resonance characteristic of B-DNA we now turn to the analysis of simulated $f_P(r)$ and $sM_P(s)$ patterns of its quasi-periodic phosphorus chains. The distribution of phosphorus atoms in the macromolecule is schematically depicted in Fig. 4. An inspection of the electron scattering pattern of the chain reveals that the only sinusoidal wave which persists at $s > 3$ Å^{-1} has the periodicity of ~ 1 Å^{-1}. With the help of a simple Fourier transform, we establish that it originates from a set of the closest phosphorus–phosphorus internuclear distances, $r_{P\ldots P} \approx 7$ Å (Fig. 4). We conclude that (i) the scattering intensity features associated with the strand-to-strand internuclear distances are damped out almost completely for $s > 3$ Å^{-1} and (ii) accounting for these features in the course of an electron diffraction experiment requires a high-resolution CCD camera. A comparison of the radial distribution functions characteristic of full DNA duplex and separate strands (Fig. 4) reveals that the strand-to-strand internuclear distances result in pronounced distance-density accumulations (spatial resonance) at $r \approx 13$, 16, and 18 Å. Because a typical hydrogen-bonded DNA base pair is about 10 Å in size, the resonant distance density accumulations at $r < 10$ Å (Fig. 3) are largely due to base pairing and stacking ($r \approx 5$ Å) in the duplex. In summary, UED study of order–disorder transitions in DNA must be feasible due to a very pronounced structural resonance (see below), but it cannot be accomplished without solid theoretical guidance.

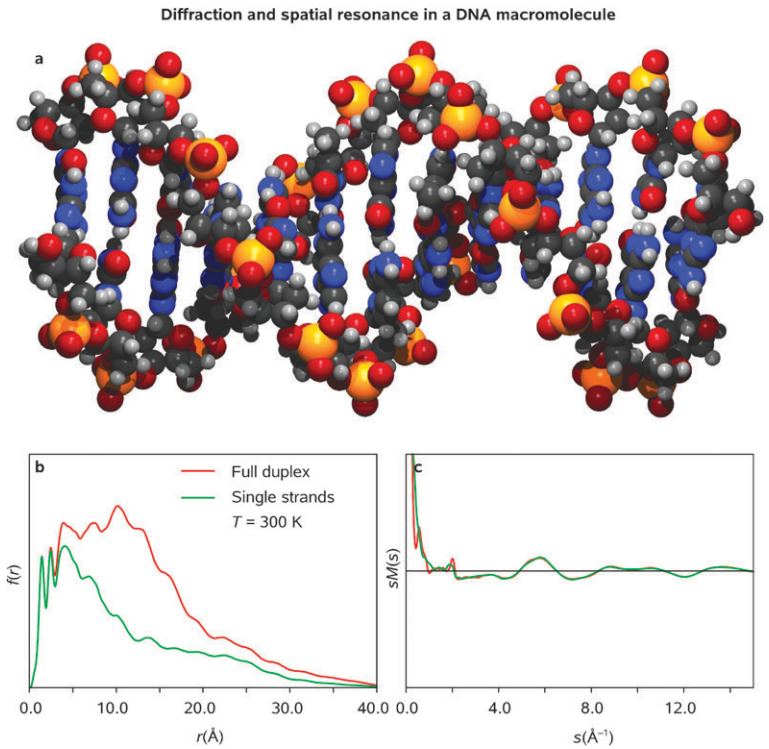

Fig. 3 Diffraction and spatial resonance in DNA. Shown is the molecular structure of a B-DNA 13-mer [5′-d(CGCGGTGTCCGCG)-3′, PDB ID: 1LAI] in aqueous solution as obtained from NMR experiments (a). Radial distribution functions ($f(r)$; b) and electron diffraction patterns ($sM(s)$; c) represent the full duplex and individual (totally uncoupled) DNA strands of the 13-mer.

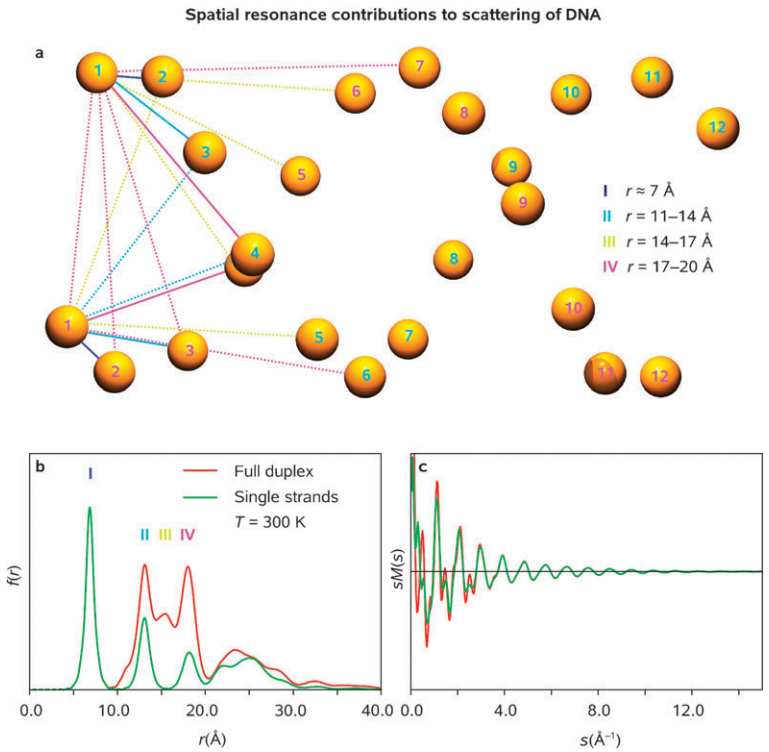

Fig. 4 Spatial resonance contributions to scattering of DNA. Shown are radial distribution functions ($f_P(r)$; b) and electron diffraction patterns ($sM_P(s)$; c) of phosphorus chains in the full duplex (a) and uncoupled DNA strands of the B-DNA 13-mer displayed in Fig. 3. Tentative assignment of the resonant domains in the radial distribution function of the duplex (numbered I to IV) was carried out in the assumption that the interstrand phosphorus–phosphorus distances propagate quasi-periodically along the duplex: $r(P_i \ldots P_{i+k}) \approx r(P_{i+m} \ldots P_{i+k+m})$.

3. Computational

The MD simulations reported here were performed on thirteen-base-pair DNA duplexes with the sequence of 5′-d(CGCGGTGTCCGCG)-3′[20] in the gas phase and in aqueous solution. Molecular construction and MD simulations were performed using the CHARMM[21] suite of programs with the all-atom CHARMM27[27] force-field parameters. The starting-point structure was obtained from the NMR experimental data deposited under 1LAI ID in the RCSB Protein Data Bank (www.pdb.org). Because DNA macromolecules are known to have an intrinsic negative charge which is responsible for their acidic character, and which is concentrated on the backbone phosphate groups, the DNA ensemble *in vacuo* will be characterized by a certain distribution of charged states due to varying degrees of macromolecular protonation. According to the electrospray experiments of ref. 13, the charged state of a typical DNA duplex *in vacuo* should be about 1/4 of its intrinsic negative charge. Since the duplexes studied here possess an inherent charge of −24, the charged states of $Q = -6$ and −12 were assumed *in vacuo* in order to assess the impact of different values of Q on the macromolecular dynamics.

Rueda *et al.*[8] manipulated the charged state of their DNA duplexes by either neutralizing selected phosphate groups or scaling down the negative charge of all phosphate groups by the appropriate fractional factor. According to the results of ref. 8, the structural stability was essentially independent of neutralization scheme. In this study, we applied the latter methodology and scaled the phosphate-group charges of the PDB structure by factors of 1/4 and 1/2 in order to obtain the desired charged states (see above). For each of the two charged states, the initial PDB structure was energy-minimized for 12 000 steps *in vacuo* and then heated to $T = 300$ K and pre-equilibrated for 100 ps. In the gas phase, the structure was further equilibrated for 400 ns at $T = 300$ K. From the latter equilibration step, $n = 200$ random DNA configurations were obtained to represent the (isolated) macromolecular ensemble at $T_0 = 300$ K. In order to assess the ensemble-averaged temperature-jump dynamics of the DNA, the above-mentioned equilibrated macromolecular ensemble was used as a starting point for three sets of $n = 200$ independent heating trajectories representing the 300 K, 600 K, and 1200 K temperature jumps. For each of these trajectories, the starting-point macromolecular configuration was heated to the final temperature, $T = T_0 + \Delta T$, within 1 ps and allowed to evolve for up to 100 ns to obtain the duplex-unfolding statistics. Because the above simulations were performed in CHARMM on isolated DNA duplexes, periodic-boundary conditions and particle-mesh Ewald (PME) method[28] were, obviously, not used.

For the solution-phase MD simulations, the original PDB structure was centered in the cubic primary-simulation cell with the initial box length of 60.5 Å, which contained 7007 TIP3P water molecules and 24 sodium atoms for a neutral system. For all solution-phase simulations, MD trajectories were obtained using periodic-boundary conditions with long-range electrostatics computed *via* the PME method.[28] Following 20 000 steps of energy minimization, the box was heated to $T_0 = 300$ K within 10 ps. Subsequently, the system was allowed to evolve for 1 ns at constant temperature and 1 atm pressure using the extended-system algorithm.[29] From this trajectory, $n = 200$ random water–DNA configurations were chosen to represent the solvated structure. Using these configurations as the starting-point structures, the temperature-jump trajectories were obtained as follows. The system was heated to $T = 600$ K within 10 ps and was maintained at constant temperature and 1 atm pressure for 6 ns to obtain the duplex-unfolding statistics (see below). The MD simulations and data analyses were then repeated for $T = 500$ K. In order to assess the fraction of intact native (Watson–Crick) base-pairing contacts as well as both local and global structural changes throughout the ensemble, two complementary types of data were collected as a function of time.

First, for all sets of independent trajectories, the fraction of each native base-pairing contact remaining intact at time t was calculated as follows. The fraction of intact hydrogen bonds was obtained for every Watson–Crick base pair and further averaged over the $n = 200$ independent trajectories to obtain the average decay of each native contact as a function of time. A hydrogen bond was defined to be 100% intact if the distance between the donated proton and the nitrogen or oxygen atom (the hydrogen acceptor) was less than 1.8 Å and the straight line joining the proton and the hydrogen acceptor was no more than 90 degrees out of the planes defined by the aromatic rings of the base pair. In addition, the smoothness of the transition between a fully intact and a fully broken hydrogen bond was enforced using an exponential attenuation of the bond strength such that the hydrogen bond would be 1/e-fold intact at a distance of 2.5 Å. These criteria are consistent with established conventions for geometry-based hydrogen bond determination,[30] and it should be noted that the fast process of base-pair disruption renders the results thus obtained to be insensitive to variance in the threshold values used.

Second, the ensemble-averaged radial distribution functions, $\langle f(r, t)\rangle_n$ and $\langle f_P(r, t)\rangle_n$, were calculated for the same t with a locally-modified version of UEDANA code using an artificial damping factor of $k = 0.02$ Å^2 to compensate for the unwanted oscillations induced by a finite data range ($s_{max} < \infty$).[26] Root-mean-square (RMS) amplitudes of thermal vibrations, $l_{ij} = u_{ij}$, were estimated using empirical equations[31] at $T = 300$ K and further extrapolated to elevated temperatures using eqn (4) of ref. 26.

Shown in Fig. 5 is the dependence of convergence behavior of both the decay of the third base-pair contact in the duplex (panels a, c, e) and the macromolecule-wide radial distribution function (panels b, d, f) as obtained from the $Q = -6$ MD simulations following a temperature jump of 1200 K *in vacuo* as a function of n. As seen in the results of the Figure for both measures of the spatiotemporal evolution in the ensemble, $\langle P_i(t)\rangle_n$ and $\langle f(r, t)\rangle_n$, averaging over $n = 100$ independent MD trajectories allows for the convergence of dynamical information into a clear picture of ensemble-level behavior. Indeed, whereas the information from single trajectories (Fig. 5a) may lead one to speculate that, for the native Watson–Crick contact $i = 3$, the base pairing may be quasi-stable for more than 20 ps after the temperature jump with the possibility of dynamically reforming the base pair at longer times, the information from averaging over one hundred trajectories

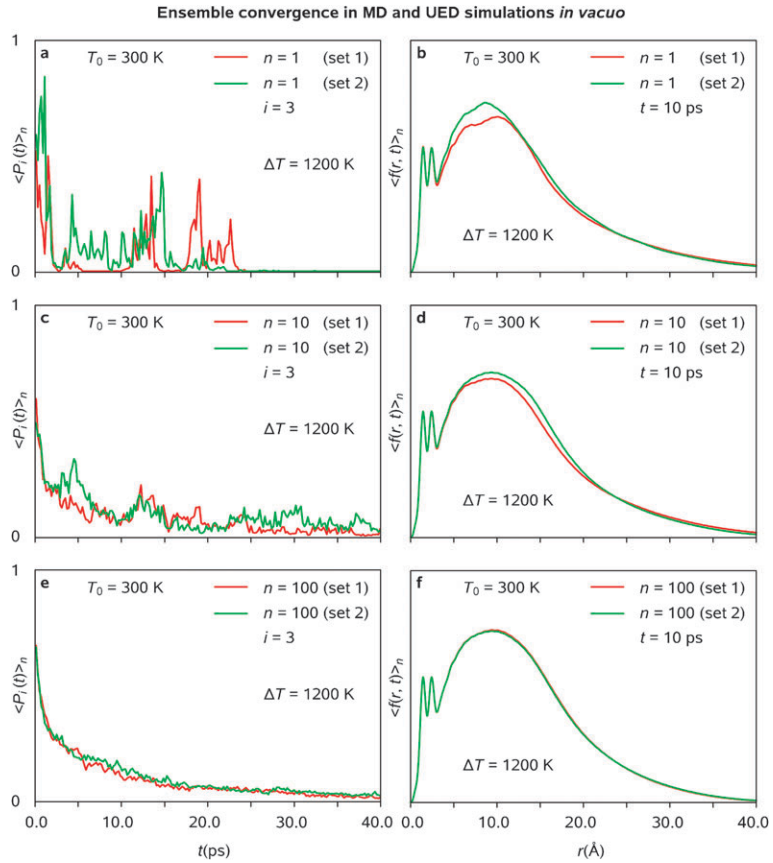

Fig. 5 Ensemble convergence in MD and UED simulations *in vacuo*. Shown is the temporal evolution of the probability density of finding the base-pair contact $i = 3$ intact upon a 1200 K temperature jump ($\langle P_i(t)\rangle_n$; a, c, e) averaged over $n = 1$, 10, and 100 MD trajectories representing two equally large macromolecular ensembles. Also presented are radial distribution functions ($\langle f(r, t)\rangle_n$; b, d, f) as obtained by averaging over $n = 1$, 10, and 100 DNA configurations representing the two ensembles at $t = 10$ ps upon the 1200 K temperature jump.

(Fig. 5e) indicates that the characteristic time scale for the base-pairing contact rupture is about 5 ps.

4. Results and discussion

Base pairing separation: unfolding trajectories

In regard to the temporal evolution of native base pairing in simulated ensembles, there are three noteworthy observations pertinent to DNA stability and (un)folding in the gas phase. First, the mechanism of DNA unzipping (or native contact rupture) turns out to be robust with respect to both the charged state assumed throughout the ensemble and the temperature jump experienced by macromolecules. Both charged states, $Q = -6$ and -12 (the latter one is omitted from this Section for the sake of conciseness) tend to undergo very similar unfolding processes at all the temperatures considered, with larger temperature jumps leading to shorter times required for complete DNA unzipping (5, 150, and 5000 ps for $\Delta T = 1200$, 600, and 300 K, respectively). Notably, the order of base-pair disruption (see Fig. 6) is preserved for all values of ΔT, whereas the unfolding time scales and behavior are virtually invariant to the charged state.

Second, DNA unzipping *in vacuo* violates the SSA, resulting in the formation of internal bulges throughout the ensemble. As outlined above, within the framework of the original KIS model the exact base-pairing configuration in DNA can be strictly defined using integers i and j which parameterize the extent of base-pair disruption on the two ends of the duplex (Fig. 2).[14b] In contrast, the gas-phase unfolding induces the formation of a "bubble" (broken base-pair sequence) in the interior of the duplex. This can be defined on the KIS landscape by an additional set of integers, i' and j', which parameterize the extent of unzipping on the two ends of the bubble, with the origin defined to be the center of the duplex. Thus, the unfolding trajectory bifurcates into the end and bubble dynamics, as indicated in Fig. 7, in which the two domains of unzipping are shown using red and green solid lines, respectively. The final step in the trajectory occurs when the red and green paths meet, with the last break in the base pairing assigned to the end dynamics by convention.

This is in contrast to the unfolding behavior of solution-phase duplexes at $\Delta T \leq 200$ K (Fig. 8), which indicates that SSA is valid in solution. However, for a somewhat higher temperature jump ($\Delta T = 300$ K, data not shown), the base pairs located at the very ends of the helices tend to break very early in the dynamics, with all the interior contacts disrupting at about the same time. Interestingly, the interior-contact unzipping does not appear to be in agreement with SSA in the latter case. It is noteworthy that, following a 300 K

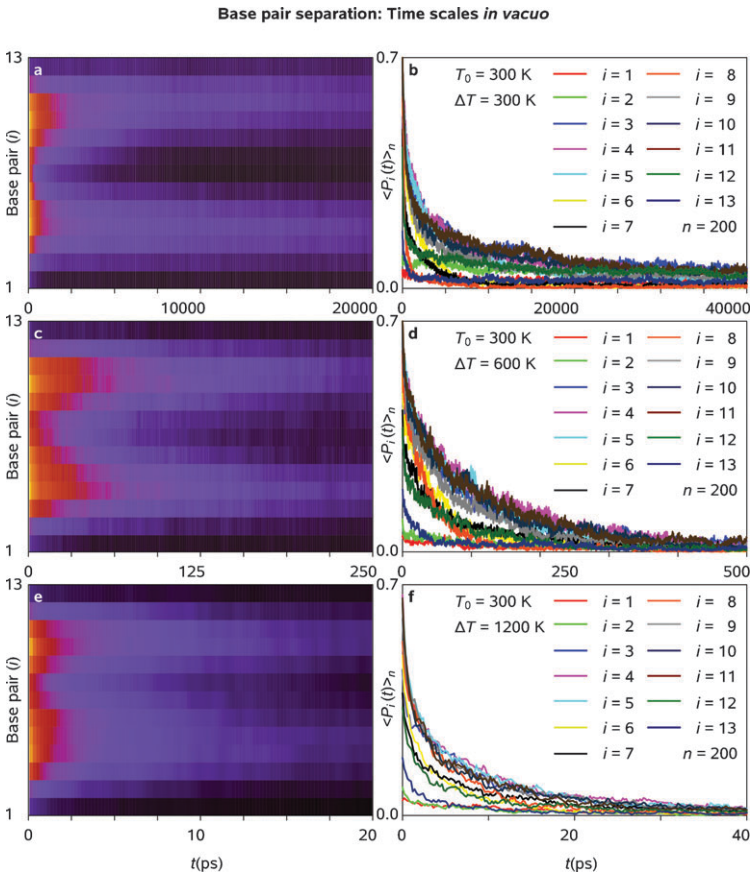

Fig. 6 Base-pair separation time scales *in vacuo*. Shown is the temporal evolution of native base pairing in DNA macromolecular ensembles upon 300 K (a, b), 600 K (c, d), and 1200 K (e, f) temperature jumps. For each Watson–Crick base-pairing contact i, $1 \leq i \leq 13$, the probability density of finding the contact intact upon the specified temperature jump at time t ($\langle P_i(t)\rangle_n$) averaged over $n = 200$ MD trajectories is plotted as a color map (yellow: 0.7, black: 0.0; a,c,e) and a set of one-dimensional profiles (b, d, f).

temperature jump in solution, the (fast) interior-contact unzipping occurs in the virtual absence of the hydration shell, the formation of which is precluded by low solvent density. Thus, the behavior we observed for $\Delta T = 300$ K in aqueous solution may represent a transition from DNA unzipping in the presence of the hydration shell to DNA unzipping *in vacuo*.

The main reason for the validity of the SSA in solution is the necessity to disrupt an extra set of stacking interactions in order to allow for the formation of an interior bubble. In the case of our duplex in the gas phase, this is compensated for by the presence of (weaker) A–T base-pairing contacts in the center of the duplex. In the solution phase, however, the hydrogen bonding with water decreases the stabilizing advantage of G–C base pairing as compared to A–T base pairing, thereby favoring unzipping from the ends of the duplex. The importance of hydrogen bonding within the duplex in the gas phase also contributes to the relatively long time required for complete strand separation, which was only observed for $\Delta T = 1200$ K. At this temperature, although hydrogen bonding of the native contacts was broken in a few picoseconds, the formation of random non-native hydrogen bonds between the two ''sticky'' strands delayed strand separation for a nanosecond despite Coulomb repulsion.

Finally, in addition to the pronounced deviation in the DNA unfolding mechanism, the presence of water which competes for hydrogen bonds with the bases and provides random thermal energy to the helices results in a striking acceleration in duplex unzipping. At 600 K, the solvated duplexes lose their native base-pairing contacts in about 100 ps as opposed to 5 nanoseconds in the gas phase. However, despite shedding light on the unzipping dynamics of base-pair disruption, the above insights do not provide a picture of local or global macromolecular contortions in response to external stress. In particular, it is of interest to know if pronounced base-pair disruption is necessary for such contortions to occur or if the DNA structure can undergo significant deformations while preserving most of its native base pairs. In the following Section we address this issue in detail both for the duplex as a whole and for the phosphate backbone of the DNA.

Conformational unfolding: radial distribution functions

Despite the rapid transition from canonical (B-DNA) to somewhat extended macromolecular structures which takes place upon DNA vaporization, the major structural motifs characteristic of hydrated DNA duplexes are largely preserved in the gas phase (Fig. 1). Regardless of the presence of hydrating environment, multiple sets of spatially coherent intramolecular distances give rise to unique resonant features in both $\langle f(r, t)\rangle_n$ and $\langle f_P(r, t)\rangle_n$ of the DNA ensembles

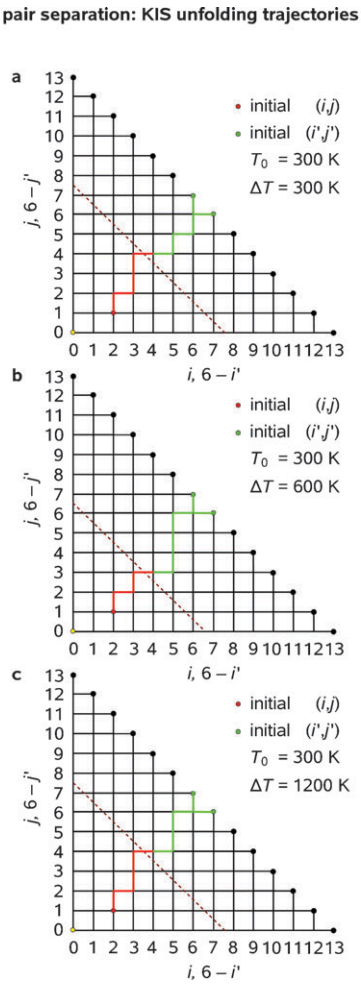

Fig. 7 KIS unfolding pathways[14b] for base-pair separation *in vacuo* as obtained by averaging over $n = 200$ MD trajectories for 300 K (a), 600 K (b), and 1200 K (c) temperature jumps. Note that because the helices are, on the average, partially unwound at $t = 0$, the starting-point configuration appears to be (2,1) rather than (0,0). Because of the formation of internal bubbles at (i',j'), see Text, the unfolding trajectories bifurcate into (i,j)- and (i',j')-related pathways which are associated with unwinding from the ends (red solid line) and from the middle of the helix (green solid line), respectively. The barrier to the unfolding transition is represented by a red dashed line. The trajectories were plotted under the assumption that a base pair becomes "uncoupled" when its ensemble-averaged survival probability decreases below 50%.

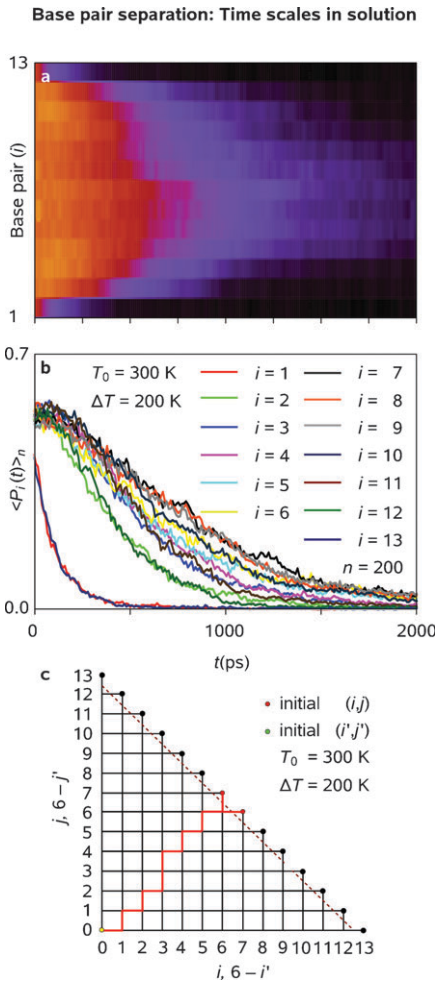

Fig. 8 Base-pair separation time scales and KIS unfolding pathways[14b] in solution. Shown is the temporal evolution of native base pairing in DNA macromolecular ensemble upon a 200 K temperature jump. For each Watson–Crick base-pairing contact i, $1 \leq i \leq 13$, the probability density of finding the contact intact upon the specified temperature jump at time t ($\langle P_i(t)\rangle_n$) averaged over $n = 200$ MD trajectories is plotted as a color map (yellow: 0.7, black: 0.0; a) and a set of one-dimensional profiles (b). The corresponding ensemble-averaged KIS unfolding trajectory is also presented (c). See caption of Fig. 7 for the criterion used in the trajectory calculation.

equilibrated at $T_0 = 300$ K ($t \equiv 0$). Thus, a comparison of radial distribution features characteristic of hydrated and isolated ensembles of 1LAI (Fig. 9) reveals that the spatial resonance associated with structural ordering in the duplexes is slightly weakened upon vaporization, which may be rationalized in terms of formation of locally ordered domains separated by somewhat distorted moieties.

However, as evidenced by sharp resonant peaks in $\langle f_P(r, 0)\rangle_n$, these changes are not associated with a pronounced loss of helicity in the backbone. It is noteworthy that the $\langle f(r)\rangle_n$ representation of isolated DNA ensembles is robust because varying the charged state of the macromolecules does not affect the major structural features of $\langle f(r)\rangle_n$ (Fig. 9). As evidenced by our MD simulations, an increased Coulomb repulsion between the strands results in somewhat decreased backbone ordering and smaller gliding shifts of stacked DNA bases which are more reminiscent of the canonical structure of B-DNA. In view of the above considerations, we limit the discussion to $Q = -6$ in this Subsection because $Q = -12$ yields virtually identical results. Importantly, deterioration of the spatial resonance in $\langle f(r, t)\rangle_n$ and $\langle f_P(r, t)\rangle_n$ upon a temperature jump ($t > 0$) provides a direct insight into the details of order–disorder transitions throughout the ensemble regardless of the actual charged state of the duplexes.

For a temperature jump of $\Delta T = 1200$ K, the complete unfolding of DNA macromolecules in the gas phase includes three distinct stages each of which is characterized by its specific time scale. First, the base-pairing and base-stacking contacts are broken throughout the duplex in the order of

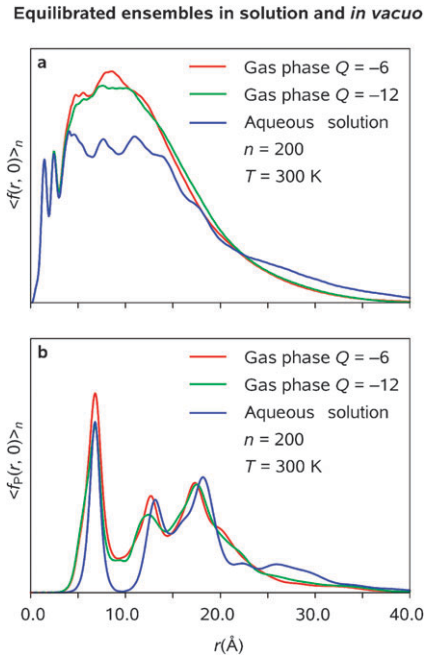

Fig. 9 Equilibrated ensembles in solution and *in vacuo*. Shown are macromolecule-wide ($\langle f(r,\ 0)\rangle_n$; a) and phosphorus chain-specific ($\langle f_P(r,\ 0)\rangle_n$; b) radial distribution functions as obtained by averaging over $n = 200$ DNA configurations representing equilibrated macromolecular ensembles in gas phase and in solution at $t = 0$ ps ($T = 300$ K). Note that changing the charged state of DNA from $Q = -6$ to $Q = -12$ does not have a pronounced effect on the features characteristic of helical resonance.

increasing energy penalty, and corresponding base pairs swing out of their "slots" in the equilibrated structure (Fig. 10). Notably, the resulting decrease of structural ordering in the duplexes, which manifests itself through gradual deterioration and smoothing of resonant features in $\langle f(r,\ t)\rangle_n$ (Fig. 11), does not affect the structure of the sugar backbone (Fig. 12). Indeed, for the initial steps of the order–disorder transition which take several ps, $\langle f_P(r,\ t)\rangle_n$ of the phosphorus chains remains virtually unchanged whereas the molecule-wide $\langle f(r,\ t)\rangle_n$ is monotonically (and quasilinearly) decreasing with time, which is also true for the resonant base-stacking feature at $r \approx 5$ Å.

Second, from $t \approx 10$ ps and on, the sugar backbones of the duplexes lose their characteristic rigidity, which leads to a partial loss of the native (helical) structure. Thus, the second-largest peak in $\langle f_P(t,\ r)\rangle_n$, $r \approx 18$ Å, starts to broaden and deteriorate, which is indicative of decreasing long-range ordering in the duplex (see Fig. 4). At $t \approx 100$ ps, the process of non-specific "coiling" of the two strands is virtually complete, and the resonant peak at $r \approx 18$ Å evolves into a residual shoulder. Simultaneously, the resonant feature at $r \approx 13$ Å transforms into a higher and broader nonresonant peak, and the nearest phosphorus–phosphorus distances $r_{P\ldots P}$ ($r \approx 7$ Å) become more spread due to the increasing randomization of the backbone structure. Third, and finally, the two DNA strands separate at $t \approx 1$ ns, which is indicative of the completion of the order–disorder transition, although the temperature has to be high enough for the strand separation to occur within a feasible simulation window.

From the results of Fig. 11 and 12, it follows that the resonance pattern in $\langle f_P(r,\ t)\rangle_n$ was largely preserved for $\Delta T < 600$ K during the course of the studied temporal intervals (the total simulation run time was chosen to be 100 and 50 ns for $\Delta T = 300$ and 600 K, respectively). Despite some minor loss of ordering at shorter times, the helical structure of the backbone remained largely intact for tens of picoseconds in the case of $\Delta T = 600$ K, and a (partial) coiling of the strands became pronounced only in hundreds of picoseconds. In contrast, the order–disorder transition at $\Delta T = 300$ K was vastly different from that discussed above for $\Delta T = 600$ and 1200 K (Fig. 11 and 12). For the 300 K temperature jump, both $\langle f(r,\ t)\rangle_n$ and $\langle f_P(r,\ t)\rangle_n$ were found to *increase* with time for shorter internuclear distances, which was indicative of ensemble-wide bending of the helices. Because the majority of base-pairing contacts remained intact at shorter time scales, the overall molecular rigidity prevented significant backbone unfolding for the first 300 ps. Moreover, even at $t = 1$ ns the helical motif was well preserved. However, the double helix undergoes significant global conformational contortions, compressing from a rod-shaped duplex to a spheroid while maintaining the majority of its initial gas-phase base pairs. Consequently, because the native contact rupture lags *behind* global conformational changes such as backbone coiling, decrease of local base-pairing and stacking is no longer isolated at the shortest times and is therefore not identifiable on the radial distribution function.

This behavior arises because base-pairing/stacking disruption and global conformational change are enthalpy- and entropy-driven processes, respectively. At lower temperatures ($\Delta T < 300$ K) conformational diffusion happens even if few disruptions occur. Since the barrier-crossing disruption process speeds up exponentially with increasing temperature while the diffusive process is weakly dependent on temperature, at a certain threshold temperature, the time scales for these two processes cross and, at sufficiently high temperatures ($\Delta T > 600$ K), all base-pairs are broken prior to the characteristic diffusion time. Crucially, this observation also suggests that UED temperature jumps be performed for $\Delta T > 300$ K to observe ultrafast local dynamics. Notably, neither 300 K nor 600 K temperature jumps were sufficient to trigger a strand separation process in the gas phase within the simulation time window.

5. Summary and conclusion

In summary, it has been demonstrated here that order–disorder transitions in gaseous DNA cannot be accounted for by a conventional "two-state" model. The folding–unfolding landscape of a free DNA macromolecule involves a number of intermediate structures which may be described as (partially) unfolded or collapsed. The time scales characteristic of native contact rupture and sugar backbone denaturation are strongly temperature dependent, so much so, in fact, that coiling of the strands may precede (at low-enough temperatures) or follow the base-pair separation. For total strand separation to occur on the nanosecond time scale, the temperature jump experienced by the ensemble has to be high enough to compensate for the "stickiness" (the intramolecular hydrogen-bonding bias)

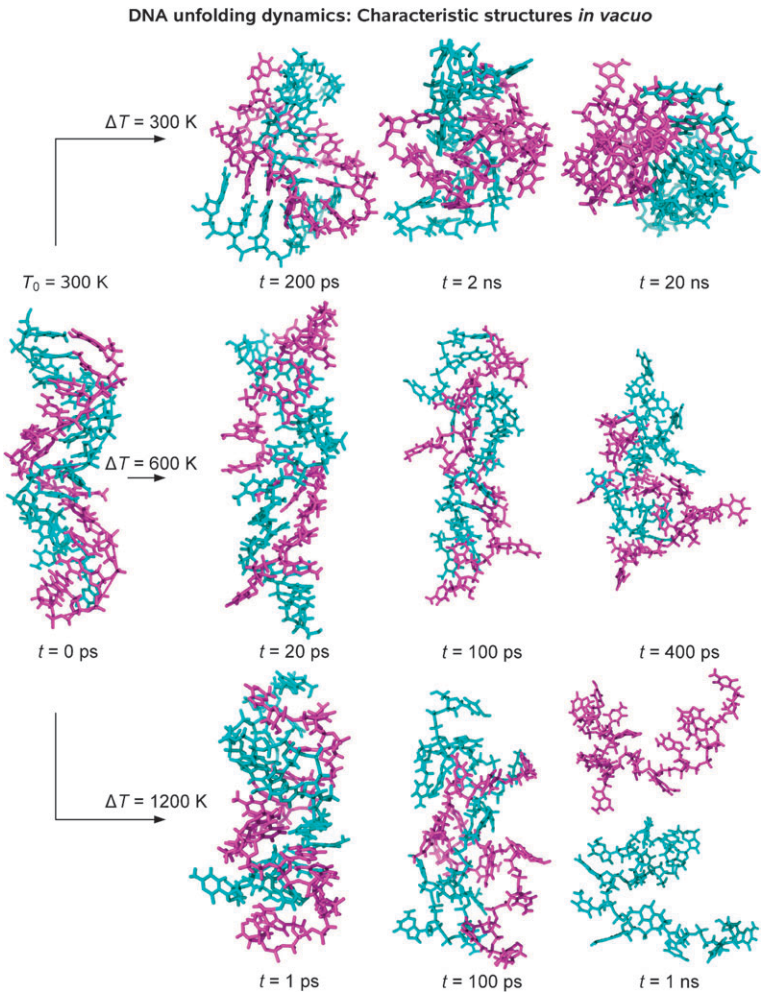

Fig. 10 DNA unfolding dynamics: characteristic structures *in vacuo*. Shown are macromolecular structures representative of different stages of DNA order–disorder transitions as obtained from MD simulations for 300 K, 600 K, and 1200 K temperature jumps. See Text.

characteristic of DNA in the gas phase. For an isolated molecular ensemble of 1LAI DNA 13-mers equilibrated at 300 K and experiencing a 1200 K temperature jump, the native contact rupture, backbone coiling, and strand separation were found to occur after 10 ps, 100 ps, and 1 ns, respectively (in the case of $Q = -12$, the strands were found to move apart somewhat faster because of a stronger Coulomb repulsion; see Fig. 13). For $\Delta T = 600$ K, the transitional behavior was qualitatively similar to that characteristic of $\Delta T = 1200$ K (apart from the unpairing/destacking and backbone coiling time scales of $\sim$60 ps and 400 ps, respectively), but for $\Delta T = 300$ K the behavior was radically different: the DNA backbone would twist and bend quasi-randomly for hundreds of picoseconds prior to any substantial base-pair disruption.

Though ensemble-averaged unfolding trajectories of free DNA macromolecules are largely insensitive to both charged-state and temperature changes, the denaturation dynamics *in vacuo* is vastly different from that found in aqueous solution. Perhaps most strikingly, the above-mentioned "stickiness" of free DNA makes order–disorder transitions in the gas phase about two orders of magnitude slower than in solution, which is caused by repetitive bonding recombination of broken base pairs in the absence of water. The (extended) double-helical structures are robust *in vacuo* because they are stable on the sub-μs time scale. From the time-dependent base-pairing analysis, all gas-phase duplexes experienced simultaneous end and interior (bubble) base-pair disruption, independent of both charged state and temperature. Thus, unlike MD simulations in solution which show that the A–T region in the center of the helix unfolds at about the same time as the neighboring (inner) base pairs, the gas-phase simulations reveal that the A–T contacts unwind almost simultaneously with the ends of the helix. Because the single-sequence approximation (SSA) is no longer valid for the isolated macromolecular ensembles, their averaged unfolding trajectories tend to display characteristic bifurcations on the KIS landscape. In the gas phase, the extra destacking energy required to form an internal bubble is compensated for by the extra stabilization of a C–G contact as compared to an A–T contact, which makes the trajectory bifurcations possible. In solution, the extra stabilization of the C–G base pairs is no longer able to compensate for destacking energies required to form a bulge because neighboring water molecules tend to hydrogen bond with the duplex.

The results reported above may suggest some biological implications. Structurally, although the double-helical motif

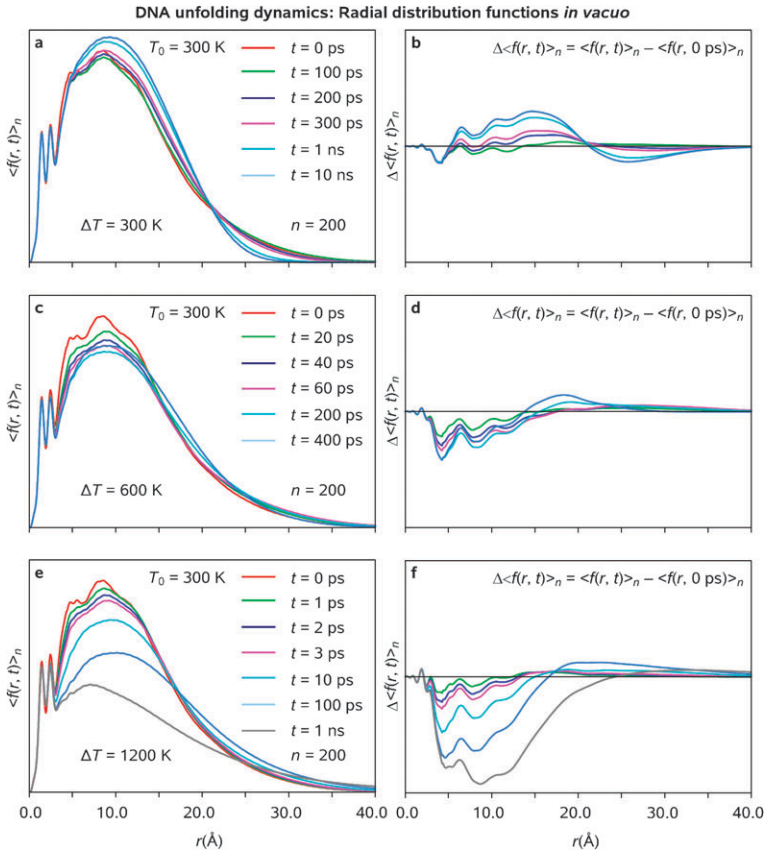

Fig. 11 DNA unfolding dynamics: radial distribution functions *in vacuo*. Shown are macromolecule-wide radial distribution functions ($\langle f(r, t)\rangle_n$; a, c, e) and corresponding temporal differences ($\Delta\langle f(r, t)\rangle_n$; b, d, f) as obtained by averaging over $n = 200$ DNA configurations at different time points t upon 300 K (a, b), 600 K (c, d), and 1200 K (e, f) temperature jumps. Note the dramatic differences in the ensemble-averaged behavior at $\Delta T = 300$ and 1200 K.

is stable on the sub-μs time scale independent of charged state, the high degree of conformational freedom and the loss of some base pairing and stacking, especially near the ends, indicate that water is necessary for the maintenance of robust double-helical structure. The contrast of the dynamical behavior in the solution and gas phases observed here suggests that the DNA-water system has fine-tuned the relative contributions of hydrogen bonding, base stacking, and entropy to achieve high-speed/low-error function. For example, the enhanced importance of intra-strand hydrogen bonding *in vacuo* led to high levels of non-native contacts and dramatically slowed the strand separation. This also led to the formation of A–T rich bubbles, which were absent in solution. The above effect may have consequences pertinent to transcription *in vivo*: since initiation of strand separation in the duplex is much less favored than propagation, this increases the chance that strand separation only occurs in response to a site-specific interaction (*e.g.* with a transcription factor) rather than thermal fluctuations. Once strand separation has been initiated, the lower relative energetic resistance to efficient unzipping allows for speedy transcription. The mechanisms and time scales reported here for the DNA double helix in a variety of solvation, charged, and temperature microensembles bring us one step closer in the quest for a complete atomic-scale picture of the effect of solvent and ligand binding on DNA stability, dynamics, and function.[32]

In conclusion, it is perhaps instructive to comment on the role of theoretical and computational modeling in the determination of ultrafast dynamics of large-scale macromolecular systems. The size and conformational flexibility of an ensemble of free DNA macromolecules represent a real challenge to both experiment and theory because the amount of individual structures to be considered is far beyond human comprehension. However, ensemble-convergent MD simulations, when aided by adequate coarse-graining approaches which allow for the exploration of the key features on the dynamical landscape without sacrificing important structural details, provide an unprecedented insight into both structure and dynamics of the macromolecular ensembles, and with state-of-the-art atomic resolution. In a similar way to ultrafast electron crystallography (UEC) experiments,[33] which utilize the spatial and temporal coherence in the specimen to reveal the structural dynamics, the numerical experiments reported here make use of both spatial resonance and ensemble averaging to extract accurate and reliable structural data at each particular point in time. Unlike computational studies which lack these features, ensemble-convergent numerical experiments can provide a much deeper and more robust

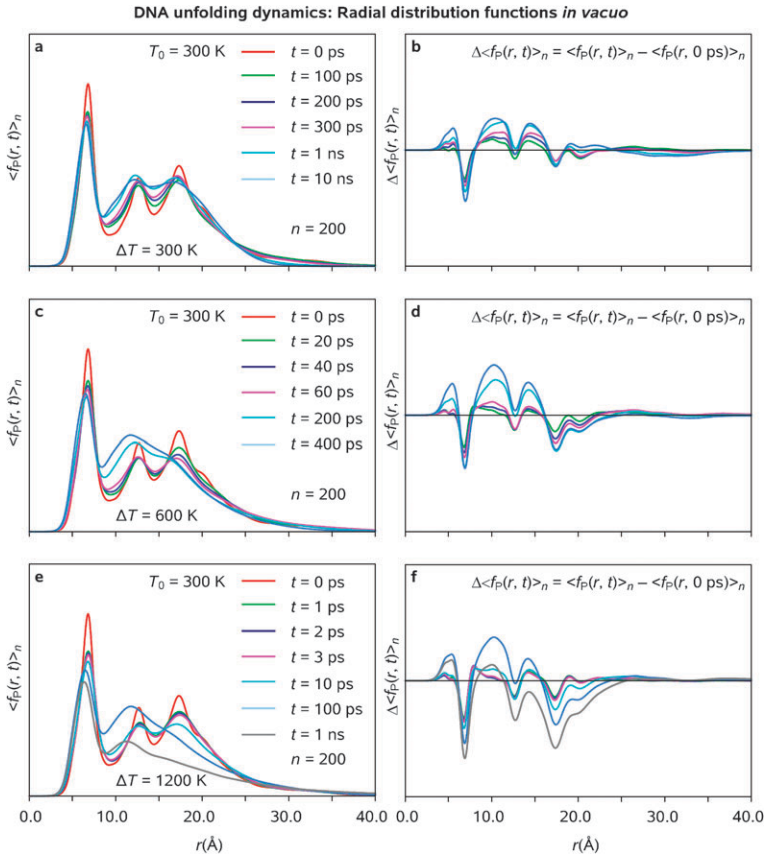

Fig. 12 DNA unfolding dynamics: radial distribution functions *in vacuo*. Shown are phosphorus-chain-specific radial distribution functions ($\langle f_P(r, t)\rangle_n$; a, c, e) and corresponding temporal differences ($\Delta\langle f_P(r, t)\rangle_n$; b, d, f) as obtained by averaging over $n = 200$ DNA configurations at different time points t upon 300 K (a, b), 600 K (c, d), and 1200 K (e, f) temperature jumps. Note the dramatic differences in the ensemble-averaged behavior at $\Delta T = 300$ and 1200 K.

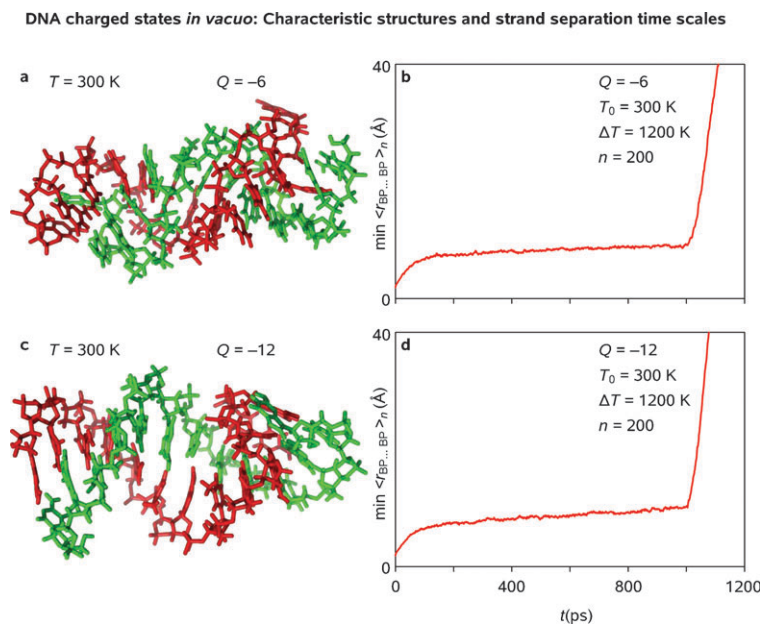

Fig. 13 DNA charged states $Q = -6$ and -12 *in vacuo*: characteristic structures (a, c) and strand separation time scales (b, d). Plotted are the minimum native base pair–base pair distances ($\min \langle r_{BP\ldots BP}\rangle_n$) as obtained by averaging over $n = 200$ DNA configurations at each particular time point t following a 1200 K temperature jump. Note that in the case of $Q = -12$ the strands tend to move apart somewhat faster because of a stronger Coulomb repulsion.

scope of detailed information on the nature of the transient behavior of DNA.

Acknowledgements

We are grateful to the National Science Foundation and National Institutes of Health (NIH grant # RO1-GM081520-01) for funding of this research. We wish to thank the referees for their thorough reviews of the paper. MML acknowledges financial support from the Krell Institute and the US Department of Energy (DoE grant # DE-FG02-97ER25308) for a graduate fellowship at Caltech.

References

1 J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 737–738.
2 (*a*) M. H. F. Wilkins, A. R. Stokes and H. R. Wilson, *Nature*, 1953, **171**, 738–740; (*b*) R. E. Franklin and R. G. Gosling, *Nature*, 1953, **171**, 740–741.
3 (*a*) A. H.-J. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. van der Marel and A. Rich, *Nature*, 1979, **282**, 680–686; (*b*) R. Wing, H. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura and R. E. Dickerson, *Nature*, 1980, **287**, 755–758.
4 D. Knorr and A. J. Sinskey, *Science*, 1985, **229**, 1224–1229.
5 K. McKeage and K. L. Goa, *Drugs*, 2001, **61**, 1599–1624.
6 See, for example: (*a*) P. W. K. Rothemund, *Nature*, 2006, **440**, 297–302; (*b*) E. S. Andersen, M. Dong, M. M. Nielsen, K. Jahn, R. Subramani, W. Mamdouh, M. M. Golas, B. Sander, H. Stark, C. L. P. Oliveira, J. S. Pedersen, V. Birkedal, F. Besenbacher, K. V. Gothelf and J. Kjems, *Nature*, 2009, **459**, 73–77; (*c*) H. Dietz, S. M. Douglas and W. M. Shih, *Science*, 2009, **325**, 725–730.
7 D. L. Hartl and E. W. Jones, *Genetics: Analysis of Genes and Genomes*, Jones & Bartlett, Sudbury, 5th edn, 2001.
8 M. Rueda, S. G. Kalko, F. J. Luque and M. Orozco, *J. Am. Chem. Soc.*, 2003, **125**, 8007–8014.
9 (*a*) T. T. Herskovits and J. P. Harrington, *Biochemistry*, 1972, **11**, 4800–4811; (*b*) L. Levine, J. A. Gordon and W. P. Jencks, *Biochemistry*, 1963, **2**, 168–175; (*c*) D. H. Turner, in *Nucleic Acids: Structure, Properties and Functions*, ed. V. A. Bloomfield, D. M. Crothers and I. Tinoco, University Science Books, Sausalito, 2000, pp. 308–310.
10 A, B, C, D and Z-structures of DNA are described in detail in: (*a*) W. Fuller, M. H. F. Wilkins, H. R. Wilson and L. D. Hamilton, *J. Mol. Biol.*, 1965, **12**, 60–76; (*b*) R. Langridge, H. R. Wilson, C. W. Hooper, M. H. F. Wilkins and L. D. Hamilton, *J. Mol. Biol.*, 1960, **2**, 19–37; (*c*) R. Langridge, D. A. Marvin, W. E. Seeds, H. R. Wilson, C. W. Hooper, M. H. F. Wilkins and L. D. Hamilton, *J. Mol. Biol.*, 1960, **2**, 38–62; (*d*) D. A. Marvin, M. Spencer, M. H. F. Wilkins and L. D. Hamilton, *J. Mol. Biol.*, 1961, **3**, 547–565; (*e*) S. Arnott, R. Chandrasekaran, L. C. Puigjaner, J. K. Walker, I. H. Hall, D. L. Birdsall and R. L. Ratliff, *Nucleic Acids Res.*, 1983, **11**, 1457–1474; (*f*) S. Arnott, R. Chandrasekaran, D. L. Birdsall, A. G. W. Leslie and R. L. Ratliff, *Nature*, 1980, **283**, 743–745.
11 T. J. Richmond and C. A. Davey, *Nature*, 2003, **423**, 145–150.
12 See, for example: (*a*) W. Fuller, T. Forsyth and A. Mahendrasingam, *Philos. Trans. R. Soc. London, Ser. B*, 2004, **359**, 1237–1248; (*b*) A. A. Lucas and P. Lambin, *Rep. Prog. Phys.*, 2005, **68**, 1181–1249 and references therein.
13 (*a*) D. C. Gale and R. D. Smith, *J. Am. Soc. Mass Spectrom.*, 1995, **6**, 1154–1164; (*b*) S. A. Hofstadler and R. H. Griffey, *Chem. Rev.*, 2001, **101**, 377–390; (*c*) V. Gabelica, E. De Pauw and F. Rosu, *J. Mass Spectrom.*, 1999, **34**, 1328–1337; (*d*) K. X. Wan, T. Shibue and M. L. Gross, *J. Am. Chem. Soc.*, 2000, **122**, 300–307; (*e*) M. L. Reyzer, J. S. Brodbelt, S. M. Kerwin and D. Kumar, *Nucleic Acids Res.*, 2001, **29**, 103e; (*f*) P. D. Schnier, J. S. Klassen, E. F. Strittmatter and E. R. Williams, *J. Am. Chem. Soc.*, 1998, **120**, 9605–9613.
14 See, for example: (*a*) H. Ma, C. Wan, A. Wu and A. H. Zewail, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 712–716; (*b*) M. M. Lin, L. Meinhold, D. Shorokhov and A. H. Zewail, *Phys. Chem. Chem. Phys.*, 2008, **10**, 4227–4239.
15 See, for example: R. M. Glaeser, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 1779–1780.
16 (*a*) M. M. Lin, D. Shorokhov and A. H. Zewail, *Chem. Phys. Lett.*, 2006, **420**, 1–7; (*b*) M. M. Lin, D. Shorokhov and A. H. Zewail, *J. Phys. Chem. A*, 2009, **113**, 4075–4093.
17 See, for example: (*a*) T. E. Cheatham III and P. A. Kollman, *Annu. Rev. Phys. Chem.*, 2000, **51**, 435–471; (*b*) J. Norberg and L. Nilsson, *Acc. Chem. Res.*, 2002, **35**, 465–472.
18 See, for example: (*a*) M. Levitt, *Cold Spring Harbor Symp. Quant. Biol.*, 1983, **47**, 251–262; (*b*) B. Tidor, K. K. Irikura, B. R. Brooks and M. Karplus, *J. Biomol. Struct. Dyn.*, 1983, **1**, 231–252; (*c*) M. Prabhakaran, S. C. Harvey, B. Mao and J. A. McCammon, *J. Biomol. Struct. Dyn.*, 1983, **1**, 357–369; (*d*) T. M. Nordlund, S. Andersson, L. Nilsson, R. Rigler, A. Graeslund and L. W. McLaughlin, *Biochemistry*, 1989, **28**, 9095–9103.
19 See, for example: (*a*) K. M. Kosikov, A. A. Gorin, V. B. Zhurkin and W. K. Olson, *J. Mol. Biol.*, 1999, **289**, 1301–1326; (*b*) J. F. Allemand, D. Bensimon, R. Lavery and V. Croquette, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 14152–14157.
20 J. P. Weisenseel, G. R. Reddy, L. J. Marnett and M. P. Stone, *Chem. Res. Toxicol.*, 2002, **15**, 127–139.
21 (*a*) B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187–217; (*b*) A. D. MacKerell, Jr., B. Brooks, C. L. Brooks, III, L. Nilsson, B. Roux, Y. Won and M. Karplus, in *The Encyclopedia of Computational Chemistry*, ed. P. v. R. Schleyer, John Wiley & Sons, Chichester, 1998, vol. 1, pp. 271–277.
22 See, for example: (*a*) E. J. Sorin, Y. M. Rhee, B. J. Nakatani and V. S. Pande, *Biophys. J.*, 2003, **85**, 790–803; (*b*) S. Kannan and M. Zacharias, *Biophys. J.*, 2007, **93**, 3218–3228; (*c*) W. Zhang and S.-J. Chen, *J. Chem. Phys.*, 2001, **114**, 7669–7681.
23 See, for example: C. Clementi, *Curr. Opin. Struct. Biol.*, 2008, **18**, 10–15.
24 SSA is a foundation of the equilibrium and kinetic zipper models frequently used in the study of helix–coil transitions in polypeptides and nucleic acids. In aqueous solution it only breaks down for very long DNA duplexes (for which there are many possible interior disruption sites). For example, Ares *et al.* demonstrated that internal bulges are only significant in aqueous solution for continuous A/T stretches of length $l = 20$ or more. See: S. Ares, N. K. Voulgarakis, K. Ø. Rasmussen and A. R. Bishop, *Phys. Rev. Lett.*, 2005, **94**, 035504.
25 S. V. Kuznetsov, Y. Shen, A. S. Benight and A. Ansari, *Biophys. J.*, 2001, **81**, 2864–2875.
26 D. Shorokhov, S. T. Park and A. H. Zewail, *ChemPhysChem*, 2005, **6**, 2228–2250.
27 (*a*) N. Foloppe and A. D. MacKerell, Jr., *J. Comput. Chem.*, 2000, **21**, 86–104; (*b*) A. D. MacKerell, Jr. and N. K. Banavali, *J. Comput. Chem.*, 2000, **21**, 105–120.
28 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
29 (*a*) H. C. Andersen, *J. Chem. Phys.*, 1980, **72**, 2384–2393; (*b*) S. Nosé and M. L. Klein, *Mol. Phys.*, 1983, **50**, 1055–1076; (*c*) W. G. Hoover, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1985, **31**, 1695–1697.
30 I. Y. Torshin, I. T. Weber and R. W. Harrison, *Protein Eng., Des. Sel.*, 2002, **15**, 359–363.
31 (*a*) V. S. Mastryukov and S. J. Cyvin, *J. Mol. Struct.*, 1975, **29**, 16–25; (*b*) E. L. Osina, V. S. Mastryukov, L. V. Vilkov and S. J. Cyvin, *J. Struct. Chem.*, 1975, **16**, 977–978; (*c*) S. J. Cyvin and V. S. Mastryukov, *J. Mol. Struct.*, 1976, **30**, 333–337; (*d*) V. S. Mastryukov and E. L. Osina, *J. Struct. Chem.*, 1976, **17**, 147–148; (*e*) V. S. Mastryukov, *J. Struct. Chem.*, 1976, **17**, 69–73; (*f*) V. S. Mastryukov, E. L. Osina, L. V. Vilkov and S. J. Cyvin, *J. Struct. Chem.*, 1976, **17**, 64–68.
32 See, for example: (*a*) S. K. Pal and A. H. Zewail, *Chem. Rev.*, 2004, **104**, 2099–2123; (*b*) P. Ball, *Chem. Rev.*, 2008, **108**, 74–108 and references therein.
33 See, for example: D.-S. Yang, N. Gedik and A. H. Zewail, *J. Phys. Chem. C*, 2007, **111**, 4889–4919.

